AD-A072 753 OKLAHOMA UNIV NORMAN DECISION PROCESSES LAB F/G 5/10
CONSISTENCY CHECKING IN HYPOTHESIS GENERATION,(U)
JUL 79 S D FISHER, C F GETTYS, C MANNING N00014-77-C-0615
UNCLASSIFIED TR-29-7-79 NL

1 OF 1
AD
A072753

END
DATE
FILMED
9 - 79
DDC

LEVEL II

12

DECISION PROCESSES LABORATORY
UNIVERSITY OF OKLAHOMA

OFFICE OF NAVAL RESEARCH
CONTRACT NUMBER N00014-77-C-0615
WORK UNIT NUMBER NR197-040

LEVEL II

12

6

Consistency Checking in Hypothesis Generation,

10

Stanley D. Fisher, Charles F. Gettys, Carol Manning,

Tom Mehle, and Suzanne Baca

14 TR-29-7-79

11 July 1979

12 42p.

prepared for

Office of Naval Research
Engineering Psychology Programs
15 Contract Number N00014-77-C-0615
Work Unit Number NR 197-040

Accession For		
NTIS GRA&I		X
DDC TAB		
Unannounced		
Justification___		
By___		
Distribution/		
Availability Codes		
Dist.	Avail and/or special	
A		

Decision Processes Laboratory
Department of Psychology
University of Oklahoma
Norman, Oklahoma 73019

Approved for public release; distribution unlimited.

410 918

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER TR 29-7-79	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) CONSISTENCY CHECKING IN HYPOTHESIS GENERATION		5. TYPE OF REPORT & PERIOD COVERED Aug. 1978-Aug. 1979
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Stanley Fisher, Charles Gettys, Carol Manning Tom Mehle, and Suzanne Baca		8. CONTRACT OR GRANT NUMBER(s) N00014-77-C-0615
9. PERFORMING ORGANIZATION NAME AND ADDRESS Decision Processes Laboratory Department of Psychology University of Oklahoma Norman, Oklahoma 73019		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 197-040
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research Engineering Psychology Programs-Code 455 800 N. Quincy Street Arlington, Virginia 22217		12. REPORT DATE 29 July 1979
		13. NUMBER OF PAGES 32
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

1) decision theory
2) hypothesis generation
3) memory search
4) semantic memory
5) memory retrieval

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Three experiments were performed to provide evidence that the generation of hypotheses in response to multiple data involves two independent cognitive processes. First, a candidate hypothesis is assumed to be retrieved or

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

Unclassified
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

(continued from block 20)

activitated in memory in response to only part of the available data. This candidate hypothesis is then assumed to be checked for consistency against the remaining data. This latter process is called "consistency checking". Experiment 1 was performed to provide evidence that consistency checking occurs during hypothesis generation. Subjects who retrieved and checked hypotheses for consistency required more time to generate a hypothesis than subjects who just retrieved hypotheses. Experiment 2 indicated that subjects performed a task analogous to the consistency checking process faster than subjects who retrieved and checked hypotheses for consistency. Experiment 3 was performed to provide evidence that consistency checking is a self-terminating process. Subjects' latencies depended upon the position of a disconfirming datum within a data set, supporting this conjecture. Although some of the predictions in experiment 1 were not supported, the results generally confirm the existence of a high-speed verification process in hypothesis generation.

Acknowledgements

This research was supported by the Office of Naval Research, Engineering Psychology Programs.

Vesta Gettys critically reviewed an early draft of this paper and delayed her vacation in the interests of a comprehensible manuscript.

Jeff Casey was also helpful in the preparation of the figure included in this report and the office staff of the University of Oklahoma Psychology Department helped in the reproduction and miscellaneous paperwork involved in the preparation of this report.

Consistency checking and Hypothesis Generation

The term "hypothesis generation" refers to the generation of answers or possible explanations to account for a given set of information which will be referred to as "data". For example, a physician generates disease hypotheses in response to a patients symptoms and the results of diagnostic tests. Gettys and Fisher (in press) and Gettys, Fisher, and Mehle (1978) have developed a tentative model of the hypothesis generation process. According to this model, hypotheses are first retrieved or activated within a semantic memory network similar to that described by Collins and Loftus (1975). Any hypothesis so retrieved is assumed to be subject to some form of plausibility assessment which may range from simple semantic verification to complex processes involving probablistic inference, depending upon task demands and the importance of the problem. The purpose of the present series of experiments is to provide evidence for the existence of a minimal form of plausibility assessment which is assumed to be intimately tied to the retrieval of hypotheses in multiple data problems.

Gettys, Fisher, and Mehle (1978) have described a study which estimated the number of activation tags a hypothesis node receives before it is retrieved as a possible hypothesis. For problems which involved three or six data, the hypotheses were estimated to be retrieved using only two or three data, respectively. These estimates suggest that some mechanism or process must exist to insure that the retrieved hypothesis is consistent with all of the available data rather than just the data from which the hypothesis was retrieved. We have named this proposed process "consistency checking" and have made three predictions concerning its operation in the overall hypothesis generation process. First, when consistency checking occurs in hypothesis generation, this process should add time required to generate a consistent hypothesis beyond that required to retrieve a hypothesis from memory. Secondly, consistency checking is expected to be a more rapid process than hypothesis generation. Thirdly, it is expected that consistency checking is a self-terminating process.

According to our theoretical analysis, the generation of hypotheses which are consistent with a set of data involves the operation of independent retrieval and consistency checking processes. The retrieval process involves activation of potential hypotheses in response to only part of the total number of data presented in a decision problem. Any retrieved hypothesis is then assumed to be checked for consistency against the remaining data. If the retrieved hypothesis is found to be consistent with the remaining data, it will be emitted as a response. However, if any datum is found to be inconsistent with the retrieved hypothesis, the entire process will be repeated until a consistent hypothesis is found.

If the consistency checking process is independent from hypothesis retrieval, it should be possible to eliminate consistency checking from hypothesis generation by instructing subjects to respond with the first retrieved hypothesis suggested by a set of data irrespective of its consistency ("first hypothesis retrieval condition"). Hypothesis generation data collected from subjects given such instructions could then be compared to data from subjects instructed to generate consistent hypotheses ("consistent hypothesis retrieval condition"). We predict that subjects instructed to generate consistent hypotheses will perform a hypothesis generation task slower than subjects instructed to respond with the first retrieved hypothesis. This time difference should be due to the occurence of the additional retrievals and consistency checking required to generate a consistent hypothesis. In addition, it predicted that this time difference will increase as a function of the number of data presented in the decision problem (data set size). This is expected to occur for two reasons. First, as data set size increases the average number of data which are checked for consistency will increase. Thus, the amount of time used for consistency checking will increase with data set size, adding disproportionately to the time needed to retrieve a hypothesis from memory. Secondly, the generation of consistent hypotheses may require the retrieval of several hypotheses, some of which will be discarded as a result of consistency checking. Since we believe that hypothesis retrieval involves only part of the available data, it is expected that the probability of generating a consistent hypothesis on the first retrieval attempt will decrease as a function of data set size. As data set size increases more retrievals will be necessary to generate a consistent hypothesis and these will add disproportionately to the time needed to retrieve a hypothesis as set size increases. For this same reason we expect the number of errors to increase as a function of set size in the "first hypothesis" retrieval condition, but not in the "consistent hypothesis" condition. Thus, we expect an interaction between these instruction conditions and data set size for both RT and error data. Regression analyses predicting hypothesis generation reaction time (RT) as a function of set size can be performed for both the "first hypothesis" and "consistent hypothesis" instruction conditions. The slopes of the best-fitting regression lines can then be compared. We predict that the slope of the "first hypothesis" retrieval instruction condition will be less than the slope of the "consistent hypothesis" retrieval instruction condition. The difference in the slopes of these lines will reflect how much time is required to perform the additional retrievals and consistency checking as set size increases and can be used as a crude estimation of the additional time required for these processes.

The second major consistency checking prediction is that it should be a more rapid process than hypothesis generation. In a semantic network model of memory (Collins and Loftus, 1975), concepts are represented as nodes interconnected by relational pathways. When a hypothesis is retrieved from such a network, activation is assumed to spread from both the nodes representing the general hypothesis category and from the data until several of these sources of activation meet at a hypothesis node. This activated node would then become a potential hypothesis. Thus, hypothesis retrieval may involve the activation of potential hypothesis nodes by relational paths leading from the data and general hypothesis category nodes. However, in consistency checking, the hypothesis is already available in memory and only the relation pathways leading to the data nodes must be activated. Thus, consistency checking is assumed to involve only the activation of relational information. From this it follows that consistency checking should occur at a faster rate than hypothesis generation.

This prediction can be examined by a task manipulation in which the time to generate a consistent hypothesis (hypothesis generation task) can be compared to the time required to perform a task which is analogous to the consistency checking process (consistency checking task). This latter task involves the initial presentation of a hypothesis, followed by the presentation of a data set. The subject then must decide if the hypothesis is consistent with the data. This task can be considered analogous to consistency checking since it eliminates the hypothesis retrieval process and involves only the verification of relational information between the hypothesis and data. We expect that the consistency checking task will be performed faster than hypothesis generation, because the hypothesis generation task requires the additional retrieval process. In addition, it is expected that the time difference between these two tasks will increase as a function of data set size since the average number of retrievals involved in the generation of consistent hypotheses is thought to follow the same function. Thus, we predict a task by data set size interaction. Regression analyses predicting hypothesis generation and consistency checking RT as a function of set size can also be performed. It is expected that the slope of the best fitting regression line for the consistency checking task will considerably less than the slope of the hypothesis generation task. The difference between these slopes can also be used as a crude estimation of the additional time needed for retrieval of a consistent hypothesis since the hypothesis generation task involves hypothesis retrieval while the consistency checking task does not.

The final prediction about consistency checking is that it is a self-terminating process. If a disconfirming relationship is found between a potential hypothesis and a single datum, the consistency checking process will stop. This prediction is based on the efficiency of this type of search. If a subject encounters a datum which is inconsistent with a hypothesis, the hypothesis is rendered implausible and it is useless to continue to verify it with the remaining data. This prediction can be examined using the same consistency checking task as was used to estimate the rate of hypothesis retrieval. If consistency checking terminates upon encountering a disconfiming hypothesis, then the latency to render a hypothesis implausible should increase as the ordinal position of the first disconfirming datum is increased in a set of data.

Three experiments were performed to verify these three major predictions. Experiment 1 involved an instructional manipulation where subjects either generated the first hypothesis suggested by a set of data or a hypothesis which was consistent with all of the same data. Experiment 2 involved a task manipulation in which subjects either generated hypotheses or checked them for consistency. Finally, Experiment 3 was designed to test the self-terminating assumption by manipulating the position of a disconfirming datum within a set of data.

Preliminary scaling study

An initial scaling study was performed to select a roughly homogeneous set of hypothesis generation problems to be used in experiments 1 and 2.

Method

Materials. A total of 100 animal hypothesis generation problems were used as materials. All problems consisted of charactersictics normally associated with different animals. The data included such items as mode of locomotion, types of appendages, native continent, food sources, color, and size. Twenty-five problems were included in each set size (1, 2, 3, and 4 data). All the data were selected by the senior author to suggest a fairly large number of animals. Thus each problem could have several correct answers.

Subjects. Twenty-four University of Oklahoma introductory psychology students served as subjects for class credit and were run in two groups of 12 subjects.

Procedure. Subjects were presented the hypothesis generation problems one at a time by an overhead projector. The order of the problems was not randomized. The one datum problems were presented first, followed successively by the two, three, and four data problems. The problems were presented on a screen and subjects were given 60 seconds to write as many animals as they could which were consistent with all the data presented on that trial. After 60 seconds the experimenter stopped the group and moved to the next problem. Subjects were given a five minute break half way through the procedure.

Analysis. The total number of generated hypotheses, the total number of unique hypotheses, and the percentage of correct and incorrect hypotheses were tabulated for each of the 100 problems used in the study. These results were used to select the problems for experiments 1 and 2.

Experiment 1

Experiment 1 was performed to provide evidence that consistency checking occurs in hypothesis generation. Experiment 1 employed an instructional manipulation designed to produce or eliminate consistency checking. In the "first hypothesis" instruction condition subjects were told to respond with the first hypothesis which was suggested by the data presented for a particular problem without concern for its plausibility. Hypotheses generated in response to these instructions should occur only as a result of the hypothesis retrieval process. This is because subjects were asked to respond with the first hypothesis which was activated by the data. Consistency checking should not have been involved since the generated hypotheses were not required to be consistent with all of the available data. In contrast, in the "consistent hypothesis" instruction condition subjects were asked to generate hypotheses which were consistent with all of the available data. This condition supposedly involves both the operation of hypothesis retrieval and consistency checking, since subjects would have to insure that any retrieved hypothesis was consistent with all of the available data. The additional time required to generate a consistent hypothesis over the amount of time required to retrieve a hypothesis can be attributed to the operation of consistency checking.

Method

Design. The basic design of experiment 1 was a two by four mixed factorial. Retrieval instructions were manipulated as a between-subjects variable with two levels (first hypothesis vs. consistent hypothesis). Data set size was manipulated as a within-subjects variable with four levels (1, 2, 3, and 4 data) and problems were nested within each level of set size. Equal numbers of male and female subjects were included within each retrieval instruction condition. Thus, sex was treated as an additional blocking variable in the design. Performance was measured by reaction time (RT) and error rates. Reaction time was defined as the time required for a subject to generate a hypothesis following data onset. Errors were measured by the correctness of the generated hypotheses in light of the data for a particular problem.

Hypothesis generation problems. Forty-eight problems were used as stimuli for the generation of hypotheses. All problems consisted of characteristics normally associated with different animals. Twelve problems of each set size (1, 2, 3, and 4 data) were selected from those presented in the preliminary scaling study on the basis of the percentage of correct responses given for a particular problem. To minimize the number of incorrect hypotheses made in the present study, the problems with the twelve highest percentage of correct responses were chosen for each set size. Overall, the percentage of correct responses ranged from 100.0 to 85.7 percent.

In addition, there were four practice problems of each set size which consisted of products and industries for which different States are noted. The products and industries were selected to suggest the names of states which had a large variety of keyboard characters in their names to familiarize subjects with the location of as many letters on the keyboard as possible before the presentation of the experimental problems.

<u>Instructions.</u> Subjects' generated hypotheses under instructions to either respond with the first hypothesis which was suggested by a set of data or to respond with a hypothesis which was consistent with all of the data for a particular problem. In the "first hypothesis" retrieval condition, subjects were told to read all of the data presented on a given problem and then respond with the first hypothesis which occured to them without regard for its correctness or plausibility. In contrast, subjects assigned to the "consistent hypothesis" retrieval condition were told to read the data presented on a given problem and then respond with a hypothesis which was consistent with all of the data. In the practice problems, a consistent hypothesis was defined as a state which was known for all of the product and industry data presented on a given trial. In the experimental problems, a consistent hypothesis was an animal name which had all of the animal characteristic data presented on a given problem. In addition, subjects in the consistent hypothesis condition were told to generate specific rather than general animal names. This was done because higher order animal classes (i.e. bird) were usually not consistent with all of the data in most of the problems.

In both instruction conditions, subjects' were told that they were being timed and were given accuracy instructions in regard to the speed-accuracy trade-off (Pachella, 1974). In the first hypothesis condition, accuracy was defined as responding with the first hypothesis which was suggested by the data, while in the consistent hypothesis condition, accuracy was defined as responding with a hypothesis which was consistent with all of the data.

Procedure. Upon entering the laboratory, subjects were seated at a Compucolor model 8001 microcomputer which presented the entire experiment except for instructions and collected all responses. First, the appropriate retrieval instructions were given in the context of the State practice problems. Once the instructions were understood by the subject, the 16 practice problems were presented in the same order for all subjects. When these were completed, the instructions were repeated in the context of the experimental animal problems and these 48 problems were presented in a random order.

Both the practice and experimental problems were presented in a similar manner. First, the data was printed in the center of the computer screen so that it could be read from left to right. At this time a software clock started in the computer. Subjects were instructed to type their responses as soon as they thought of an appropriate hypothesis, and the first keystrike of their response stopped the clock and measured the latency of hypothesis generation. All subjects were forced to give an answer to all problems. Once the entire response had been typed and any spelling errors corrected, the subject pressed the "shift" key to advance the program to the next problem. This had the effect of erasing the screen and producing a 1.5 second delay before the next series of data was presented. At no time did subjects receive feedback concerning the correctness of their responses.

Subjects. Forty-eight University of Oklahoma introductory psychology students served as subjects. Limited typing skills were required of all subjects who participated. Subjects were randomly assigned to one of the retrieval conditions upon entering the laboratory. The data of an additional 17 subjects were discarded because of equipment failure, the inability to type, or because the subject was not a native speaker of English.

Results and Discussion

An ANOVA was performed on the trimmed means for each set size of a subjects latency data. Any individual RT was excluded from these means if it was above or below .75 standard deviations from the mean of all RTs within that set size. This cutoff criterion was chosen because it usually eliminated only extreme outlying latencies. This was necessary because subjects occasionally "drew a blank" and had latencies longer than two minutes. The analysis resulted in the expected main effects of instructions, $F(1,44) = 8.48$, $MSe = 10.38$, $p < .01$, and set size, $F(3,132) = 147.79$, $MSe = 10.38$, $p < .001$. However, the instructions by set size interaction was not significant. The means of both instruction conditions for each set size are presented in Table 1.

Insert Table 1 about here

Another ANOVA was performed on the means of the correctly answered problems of each set size. These means were used so the results of the present experiment could be compared to equivalent data obtained in experiment 2. Again, there were significant main effects of instructions, $F(1,44) = 4.95$, $MSe = 31.19$, $p < .05$, and set size, $F(3,132) = 47.94$, $MSe = 5.99$, $p < .001$, and the instruction by set size interaction was not significant. Regression analyses were performed

TABLE 1

TRIMMED MEAN REACTION TIME IN SECONDS AS A FUNCTION OF INSTRUCTIONS AND SET SIZE IN EXPERIMENT 1

INSTRUCTIONS	SET SIZE 1	2	3	4
FIRST HYPOTHESIS	3.13	3.92	6.76	6.38
CONSISTENT HYPOTHESIS	4.08	5.43	8.97	8.26

on these means predicting RT as a function of set size. The resulting correlations were .905 and .861 for the first and consistent hypothesis conditions, respectively. The slopes obtained from these analyses were then used to estimate the extra time required by consistency checking. The slope obtained for the "first hypothesis" condition was 1.49 second/datum while the slope of the "consistent hypothesis" condition was 1.85 second/datum. The difference, .36 second/datum, is an estimate of the extra time which the consistency checking process requires. The means of the correctly answered problems for both instruction conditions and all set sizes are presented in Table 2.

Insert Table 2 about here

These means are also presented graphically along with the best-fitting regression lines in Figure 1a.

Insert Figure 1 about here

The significant set size effect is consistent with the results of Graesser and Mandler (1978) who found that the time required to generate the name of a dimension which is common to a set of words increases as a function of set

TABLE 2
MEAN REACTION TIME IN SECONDS OF CORRECT HYPOTHESES
AS A FUNCTION OF INSTRUCTIONS AND SET SIZE IN EXPERIMENT 1

INSTRUCTIONS	SET SIZE 1	2	3	4
FIRST HYPOTHESIS	3.42	4.41	7.75	7.28
CONSISTENT HYPOTHESIS	4.74	5.57	10.42	9.30

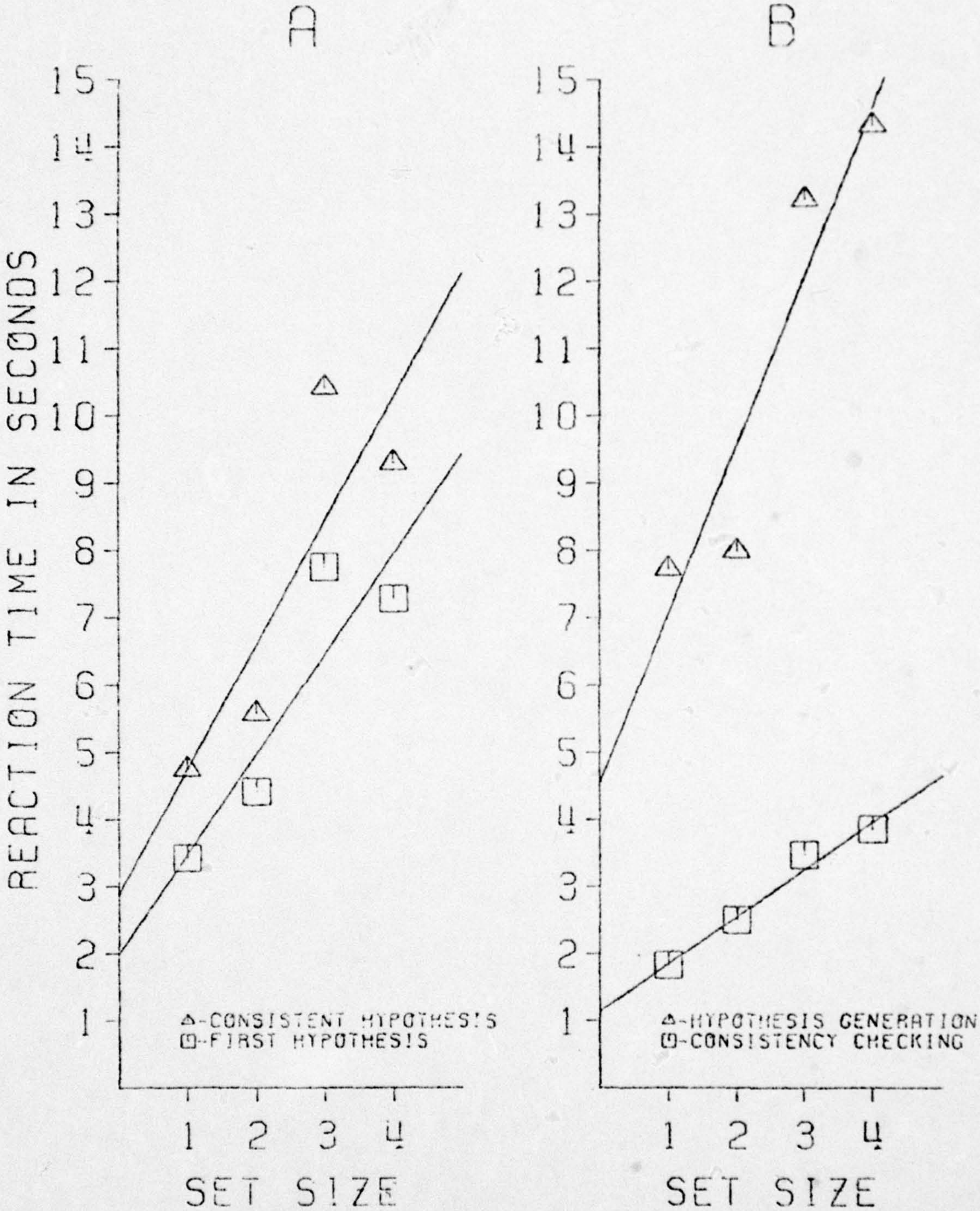

Figure 1. Mean reaction time of correctly answered problems. A) Experiment 1 for the "first" and "consistent" instruction conditions, and B) Experiment 2 for the hypothesis generation and consistency checking tasks. Also shown are the best-fitting regression lines for both instruction and task conditions.

size. However, the present results show that RT does not increase monotonically with set size since the mean of set size four is actually lower than that of set size three for both instruction conditions. This was probably due to the specific three-data problems used in the experiment. Our conjecture about this finding is that the three-data problems were generally more difficult for our subjects than the four-data problems. This was an unfortunate result of not being able to use randomly selected data for each subject. Fixed data sets were used to avoid the possibility that the same data would be presented to the same subject twice and also because some random selections of animal characteristics would have no correct answers (i.e. has wings, has four legs), or have correct answers which a typical subject would not know.

The significant instruction effect supports the prediction that consistency checking occurs during the hypothesis retrieval process. However, the failure to find a significant instructions by set size interaction is not consistent with the prediction that disproportionately more retrievals and thus more consistency checking will occur as set size increases. However, another result was obtained which is consistent with our prediction; the analysis performed on the number of errors made within each set size condition per subject resulted in a significant set size effect, $F(3,132) = 8.15$, $MSe = 1.06$, $p < .001$. The mean number of errors for set sizes 1 through 4 were 1.25, 1.75, 1.79, and 2.29, respectively. However, the instructions by set size interaction was not significant.

This result is consistent with the idea that hypotheses are retrieved using only part of the data. If hypothesis retrieval is based upon all of the data, irregardless of set size, then a constant error rate would be expected across all set sizes. However, if retrieval is based on only part of the data then it would be expected that errors should increase as a function of set size. Gettys, Fisher, and Mehle (1978) estimated that the number of data from which a hypothesis is retrieved increases disproportionately slower than data set size. This means that the probability of a retrieved hypothesis being consistent with all of the data will become smaller as data set size increases. Since the probability of an error is a positive function of the retrieved hypothesis being inconsistent with part of the data, the number of errors should increase with set size.

In addition, more errors were made by the first hypothesis condition (1.99) than by the consistent hypothesis condition (1.55), but this difference did not attain traditional levels of significance, $F(1,44) = 2.96$, $MSe = 3.09$, $p < .10$. This trend in the means also is consistent with our prediction; subjects who do not consistency check should generate more hypotheses which are inconsistent with all of the data. The error data is consistent with the prediction that "first hypothesis" condition subjects would have more errors as set size increased, but is not consistent with the prediction that "consistent hypothesis" subjects would have a constant number of errors across set size. Evidently, subjects in the "consistent hypothesis" condition did not always check their answers for consistency, or did not have enough knowledge to generate consistent hypotheses.

The overall low error rate between the "first" and "consistent" instruction conditions also indicates that subjects in the "first hypothesis" condition probably engaged in some plausibility assessment before emitting their responses. Thus, the instruction manipulation was not completely successful in either producing or eliminating the consistency checking process. The failure to find a significant interaction between instructions and set size with the RT data was also probably the result of the weakness of the instruction manipulation.

Examination of possible artifacts. Another possible explanation of the instruction effect is that subjects in the "first hypothesis" condition tended to repeat answers more often than in the "consistent hypothesis" condition. Collins and Loftus (1975) predict that a previously activated concept will be relatively easier to reactivate than a non-activated concept. This would lead to the prediction that a previously generated hypothesis would require less time to generate than a hypothesis which has been retrieved for the first time. Loftus and Loftus (1974) and Loftus (1973) have found that repeated retrievals from the a category result in faster latencies than the first retrieval from the same category. To test this prediction, the number of repeated hypotheses were found for each subject and entered into an ANOVA. The results showed that females (13.37) made significantly more repetitions than males (10.54), $F(1,44) = 5.01$, $MSe = 19.2$, $p < .05$, but the instruction effect was not significant, suggesting that response repetition does not account for the observed instructions effect.

Another possible explanation for the instruction effect is that the instructions given to the "consistent hypothesis" subjects asked for specific animal names rather than general categories of animals while no such restriction was imposed on the "first hypothesis" subjects. The retrieval of specific animal names may involve a more extensive memory search than the retrieval of general categories of animal (e.g. bird, fish, snake, etc.). If this is true then it would be expected that the "consistent hypothesis" condition would be slower than the "first hypothesis" condition. First, the number of general animal names given by each subject in both instruction conditions were counted. An ANOVA performed on these data resulted in a significant effect of instructions, $F(1,44) = 10.64$, $MSe = 9.05$, $p < .01$. The mean number of general category names for the "first and "consistent" hypothesis conditions were 8.25 and 5.42, respectively. In addition, males (5.58) produced significantly fewer general names than females (8.08), $F(1,44) = 8.28$, $MSe = 9.05$, $p < .01$. If the instruction effect on RT was due to this difference in the number of general names, then it would be expected that there would be a high correlation between the number of general names emitted and the mean RT for each subject. However, the correlation between these measures was -.210 which is not convincing evidence for this argument.

Therefore, the instruction effect can tentatively be explained by the operation of consistency checking process, although the failure to find a significant instructions by set size interaction was disquieting. As previously mentioned this was probably due to the weak effect of instructions. In addition, there may have not been large enough data set sizes in the present experiment for the interaction to be manifest. In spite of this failure, the instruction manipulation did produce results predicted by the operation of consistency checking in hypothesis generation.

Experiment 2

Experiment 2 was designed to demonstrate that consistency checking is a more rapid process than hypothesis generation. This was done by comparing a hypothesis generation task to a consistency checking task. The hypothesis generation task was identical to the "consistent hypothesis" instruction condition in experiment 1. Again it was assumed that this condition involved the operation of both hypothesis retrieval and consistency checking as discussed previously. The consistency checking task involved presenting a hypothesis followed by the presentation of data. The presentation of a hypothesis prior to the data eliminated the hypothesis retrieval process from this task. After the data was presented the subject checked the data and hypothesis for consistency and then made either a "yes" or "no" response depending upon whether the all of the data was consistent with the hypothesis. This "consistency checking" task is analogous to the proposed consistency checking process for two reasons. First, the presentation of the hypothesis eliminates the hypothesis retrieval process, thus simulating the situation where a hypothesis has been activated by other data. Secondly, the task involves only the semantic verification of semantic relationships between the hypothesis and data.

Method

Design. Experiment 2 was a two by four mixed factorial design in which task (hypothesis generation vs. consistency checking) was a between-subjects variable and data set size (1, 2, 3, or 4 data) was a within-subjects variable.

This design was a "group-yoked" design. This design was termed "group-yoked" because the hypotheses generated by the subjects in the hypothesis generation task were presented to subjects in the consistency checking task. Thus the subjects in the consistency checking task were yoked to the *correct* responses given by subjects in the hypothesis generation task. This yoking could not be done at the individual level because a typical hypothesis generation subject does not always give correct answers to the problems. To eliminate this problem, all of the correct answers for a particular hypothesis generation problem were pooled and presented to consistency checking subjects in the same proportions as they were originally generated. Equal numbers of male and female subjects were included in both task conditions and served as an additional blocking variable in the design. Performance was measured by the latency to either generate a hypothesis or check a hypothesis against a set of data in the hypothesis generation and consistency checking tasks, respectively.

Subjects. Forty-eight University of Oklahoma introductory psychology students served as subjects for class credit. The data of an additional 20 subjects were discarded beacause of equipment failure, poor typing skills, or because the subject was not a native speaker of English.

Materials. The same 48 hypothesis generation problems used in experiment 1 were also used in the present experiment. In the consistency checking task there were 48 "true" and 48 "false" problems. The same animal characteristic data used in experiment 1 were also used in the "true" consistency checking problems. The hypotheses presented in the "true" consistency checking problems were the correct hypotheses generated by subjects who completed the hypothesis generation task using the same data. Different correct hypotheses generated by the hypothesis generation subjects were presented to consistency checking subjects in approximately the same proportions as they had been emitted in the hypothesis generation task.

An additional 48 "false" consistency checking problems were constructed to make 'yes' and 'no' responses equally probable in the consistency checking task. Both the data and hypotheses used in these "false" problems were selected by the experimenter. Twelve "false" problems were used for each data set size and one datum was chosen to be inconsistent with the hypothesis used in each problem. In the case of multiple data problems the position of the disconfirming datum was counterbalanced across different problems.

Procedure. Since the hypotheses which were checked for consistency were generated by subjects in the hypothesis generation task, it was necessary to run the hypothesis generation task before the consistency checking task. The procedure used in the hypothesis generation task was identical to that used in the consistent hypothesis retrieval condition in experiment 1.

Subjects in the consistency checking task were seated at a Compucolor model 8001 microcomputer which presented the entire experiment except for instructions and also recorded all responses. Then 16 practice problems were presented which involved the same data as used in experiment 1, followed by the 96 experimental animal problems. In both the practice experimental problems, subjects were told to make a 'yes' response if the presented hypothesis was consistent with all the data or a 'no' response if any one datum was inconsistent with the hypothesis.

A consistency checking trial began with the presentation of a hypothesis which was a State name in the practice problems and an animal name in the experiemental problems. This hypothesis remained on the screen until the subject pressed the 'space bar' on the computer keyboard. At that time the hypothesis was erased, followed by a 1.5 second delay and the presentation of the data. The subject then pressed either the "Z" or "/" key to indicate either a "yes" or "no" response and the position of the "yes" and "no" keys were counterbalanced across subjects. These keys were located on the bottom row of the computer keyboard and were chosen as response keys because they were widely separated and subjects could easily keep their fingers poised above these two buttons. When a response was made, the software clock stopped, and the response was printed on the screen beneath the data and remained there for a 1 second interval. Finally, the screen was cleared and the next hypothesis was presented. This procedure was repeated until all of the problems had been presented.

Results and Discussion

An ANOVA was performed using the mean latencies of the correctly answered hypothesis generation and "true" consistency checking problems. Both main effects of task, $F(1,44) = 169.61$, $MSe = 17.64$, $p < .0001$, and set size, $F(1,132) = 64.84$, $MSe = 3.47$, $p < .0001$, were significant. In addition, the task by set size interaction was significant, $F(3,132) = 22.81$, $MSe = 3.47$, $p < .0001$. Regression analyses performed on the two task conditions predicting RT as a function of set size produced correlations of .938 and .988 for the hypothesis generation and consistency checking tasks, respectively. The slopes of the best-fitting lines obtained from these analyses were 2.499 and .703 for the hypothesis generation and consistency checking tasks, respectively. The difference between these two slopes was 1.796. The mean RT's for both task conditions across set size are presented in Table 3.

Insert Table 3 about here

The same mean RT's along with their respective best-fitting regression lines are presented in Figure 1b.

By comparing Figures 1a and 1b, it can be seen that the y-intercept is noticably greater for the hypothesis generation task obtained in experiment 2 than for the identical "consistent hypothesis" condition in experiment 1. This difference may have been due to a difference in typing skills between the subjects in the two experiments. Subjects were allowed to sign up for experiment 2 without having typing skills and only those subjects which had a great difficulty in finding key locations were discarded. However, in experiment 1, subjects were not allowed to sign up unless they were familiar with a typewriter keyboard.

The significant task effect demonstrates that a consistentcy checking task is performed more rapidly than a hypothesis generation task. The large difference

TABLE 3
MEAN REACTION TIME IN SECONDS OF CORRECT HYPOTHESES
AS A FUNCTION OF TASK AND SET SIZE IN EXPERIMENT 2

TASK	SET SIZE 1	2	3	4
CONSISTENCY CHECKING	1.83	2.50	3.47	3.85
HYPOTHESIS GENERATION	7.72	7.99	13.21	14.31

in the y-intercepts between the tasks is due to the different responses made by the two groups. The hypothesis generation task involved many different keys on the keyboard, while the consistency checking task involved only two. Despite this difference in y-intercepts, it can be seen that the slope of the best fitting regression line of the consistency checking task is considerably lower than that of the hypothesis generation task. This indicates that the amount of time needed to process one additional datum was considerably greater for the hypothesis generation task than the consistency checking task. This result provides support for the prediction that consistency checking is a more rapid process than hypothesis generation. In addition, the crude estimate of the additional time required for consistency checking obtained in experiment 1 (.361 second/datum) is also considerably less than the estimate of the additional time required for hypothesis retrieval obtained in the present experiment (1.798 second/datum). This difference in estimates is also consistent with the prediction that consistency checking is a high-speed verification process rather than a memory search process.

Experiment 3

Experiment 3 was performed to test the prediction that consistency checking is a self-terminating process. This was done by varying the position of a disconfirming datum within three data consistency checking problems similar to those used in experiment 2.

Method

Design. Experiment 3 was a within-subjects design where the independent variable was the ordinal position of a disconfirming datum within a series of three data (position 1, 2, 3 or no disconfirming datum). Performance was measured by the latency to determine whether a hypothesis was consistent or inconsistent with all of the available data. Sex and "yes-no" key positions were also included in the design as a blocking and counterbalancing variables, respectively. In addition the position of the disconfirming datum was counterbalanced for a given problem across subjects.

Subjects. Twenty-four University of Oklahoma introductory psychology students served as subjects for class credit. All were randomly assigned to the counterbalancing conditions.

Materials. Eighteen practice and sixty experimental consistency checking problems served as materials. All consisted of a hypothesis and three data. The practice problems involved checking countries against products and industries, occupations against tools, and animals against characteristics. The experimental problems only involved checking occupations against tools and animals against characteristics. Within each problem type there were 15 problems in which the hypothesis was consistent with all the data and 15 problems where one datum was inconsistent with the hypothesis. Within these disconfirming problems, there were five problems with the disconfiming datum in the first, second, and third positions.

Procedure. Upon entering the laboratory, subjects were seated at a Compucolor model 8001 microcomputer which presented the entire experiment and recorded all responses. First, instructions were given about the nature of the consistency checking problems and the the practice problems were presented. Then the instructions were repeated and followed by the experimental problems. Both the practice and experimental problems were presented in the same order for all subjects. The procedure and instructions used in experiment 3 were similar to those used in the consistency checking task used in experiment 2. First, a hypothesis was presented on the screen and remained there until the subject pressed the "space bar". This erased the hypothesis and following a 1 second delay the three data were presented in a vertical list so that they could be read from top to bottom. The data remained on the screen until the subject pressed either the "Z" or "/" key on the keyboard, depending upon whether their response was "yes" or "no" and which of these two keys represented these responses. Then the response was printed on the screen and remained there for 1 second, after which, the screen was cleared and the next hypothesis was presented.

Results and Discussion

A within-subjects ANOVA was performed on the trimmed means obtained for each type of disconfirming datum problem and the confirming problems for each subject. Any individual RT was discarded if it was above or below .75 standard deviations from the mean of the distribution of all the RT's for a particular problem. The results of this analysis indicated a significant effect of the position of the disconfirming datum, $F(3,60) = 10.91$, $MSe = 1948.13$, $p < .001$. The means of the first, second and third disconfirming datum positions were 2.78, 3.23, and 3.49 seconds, respectively, while the mean of the confirming or true problems was 3.14 seconds. Tukey pairwise comparisons indicated that position 2 and 3 problems were disconfirmed significantly slower than position

1 problems, $t(1,60) = 3.53$ and 5.57, respectively, error term $= 12.71$, $p < .05$, but position 2 problems were not disconfirmed significantly faster than position 3 problems. Thus, as predicted, RT increased as a function of the position of the disconfirming datum. This result is consistent with the prediction that consistency checking is self-terminating. If consistency checking were an exhaustive process then subjects should continue checking a hypothesis after encountering a disconfirming datum. However, the present results suggest that subjects stop consistency checking when a disconfirming relationship is found between a datum and a hypothesis. The nonsignificant difference between the position 2 and 3 problems, however, suggest that some subjects did tend to read the last datum, but evidently most subjects stopped reading if the disconfirming datum was in the first position. Also of interest was a regression analysis performed predicting RT as a function of the position of the disconfirming datum. The slope of the best fitting line was .35 second/datum which is remarkably close to the .36 second/datum estimate of the additional time required for consistency checking obtained in experiment 1. This result provides converging evidence that consistency checking is a more rapid process than hypothesis retrieval.

Summary

In summary, the results of experiment 1 demonstrated that subjects who retrieved and checked hypotheses for consistency required more time to generate hypotheses than subjects who just retrieved hypotheses. This finding provides evidence that consistency checking occurs in the hypothesis generation process.

However, the predicted interaction between instructions and set size was not found. We believe this failure was the result of an ineffective instructional manipulation. Evidently, subjects in the "first hypothesis" condition inadvertantly checked their responses for consistency since they did not make significantly more errors than subjects in the "consistent hypothesis" condition. In addition, subjects in the "consistent hypothesis" condition did not always produce consistent hypotheses, especially in the larger set sizes. However, a similar interaction was found in experiment 2 which involved a task manipulation. This interaction demonstrated that consistency checking is performed much more rapidly than hypothesis generation. Finally, experiment 3 provided evidence that consistency checking is a self-terminating process.

References

Collins, A. M. and Loftus, E. F. A spreading-activation theory of semantic processing. *Psychological Review*, 1975, *82*, 407-428.

Gettys, C. and Fisher, S. D. Hypothesis plausibility and hypothesis generation. *Organizational Behavior and Human Performance*, in press.

Gettys, C., Fisher, S. D., and Mehle, T. *Hypothesis generation and plausibility assessment.* (Annual Report TR 15-10-78). Oklahoma: University of Oklahoma, Decision Processes Laboratory, October 1978. (Can be obtained from the Defense Documentation Center ADA060 787.)

Graesser, A. and Mandler, G. Limited processing capacity restrains the storage of unrelated sets of words and retrieval from natural categories. *Journal of Experimental Psychology: Human Learning and Memory*, 1978, *4*, 86-100.

Loftus, E. F. Activation of semantic memory, *Journal of American Psychology*, 1973, *86*, 331-337.

Loftus, G. R. and Loftus, E. F. The influence of one memory retrieval on a subsequent memory retrieval. *Memory and Cognition*, 1974, *2*, 467-471.

Pachella, R. G. The interpretation of reaction time in information-processing research. In B. H. Kantowitz (Ed.), *Human Information Processing: Tutorials in Performance and Cognition.* Hillsdale, N. J., Lawrence Erlbaum, 1974.

OFFICE OF NAVAL RESEARCH, CODE 455
TECHNICAL REPORTS DISTRIBUTION LIST

CDR Paul R. Chatelier
Military Assistant for Training and
Personnel Technology
Office of the Deputy Under Secretary
of Defense
OUSDRE (E&LS)
Pentagon, Room 3D129
Washington, D.C. 20301

Director
Engineering Psychology Programs
Code 455
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217 (5 cys)

Director
Analysis and Support Division
Code 230
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Director
Naval Analysis Programs
Code 431
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Director
Operations Research Programs
Code 434
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Director
Statistics and Probability Program
Code 436
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Director
Information Systems Program
Code 437
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Special Assistant for Marine
Corps Matters
Code 100M
Office of Naval Research
800 North Quincy Street
Arlington, Virginia 22217

Commanding Officer
ONR Branch Office
ATTN: Dr. J. Lester
Building 114, Section D
666 Summer Street
Boston, Massachusetts 02210

Commanding Officer
ONR Branch Office
ATTN: Dr. C. Davis
536 South Clark Street
Chicago, Illinois 60605

Commanding Officer
ONR Branch Office
ATTN: Dr. E. Gloye
1030 East Green Street
Pasadena, California 91106

Office of Naval Research
Scientific Liaison Group
American Embassy, Room A-407
APO San Francisco, California 96503

Director
Naval Research Laboratory
Technical Information Division
Code 2627
Washington, D.C. 20375 (6 cys)

ONR Code 455, Technical Reports Distribution List

Dr. Ronald Howard
Department of Engineering-Economic Systems
Stanford University
Stanford, California 94305

Dr. William Howell
Department of Psychology
Rice University
Houston, Texas 77001

Journal Supplement Abstract Service
American Psychological Association
1200 17th Street, N.W.
Washington, D.C. 20036 (3 cys)

Dr. Clinton Kelly
Decisions and Designs, Inc.
8400 Westpark Drive, Suite 600
P.O. Box 907
McLean, Virginia 22101

Dr. Robert R. Mackie
Human Factors Research, Inc.
5775 Dawson Avenue
Goleta, California 93017

Dr. Gary McClelland
Institute of Behavioral Sciences
University of Colorado
Boulder, Colorado 80309

Human Resources Research Office
300 N. Washington Street
Alexandria, Virginia 22314

Dr. Miley Merkhofer
Stanford Research Institute
Decision Analysis Group
Menlo Park, California 94025

Dr. Terence R. Mitchell
University of Washington
Seattle, Washington 98195

Dr. Jesse Orlansky
Institute for Defense Analyses
400 Army-Navy Drive
Arlington, Virginia 22202

Professor Judea Pearl
Engineering Systems Department
University of California-Los Angeles
405 Hilgard Avenue
Los Angeles, California 90024

Professor Howard Raiffa
Graduate School of Business Administration
Harvard University
Soldiers Field Road
Boston, Massachusetts 02163

Dr. Paul Slovic
Decision Research
1201 Oak Street
Eugene, Oregon 97401

Dr. J. A. Swets
Bolt, Beranek & Newman, Inc.
50 Moulton Street
Cambridge, California 94305

Dr. W. S. Vaughan
Oceanautics, Inc.
422 6th Street
Annapolis, Maryland 21403

Dr. Gershon Weltman
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, California 91364

Dr. David Dianich
Chairman, Dept. of Business and Economics
Salisbury State College
Salisbury, Maryland 21801

ONR, Code 455, Technical Reports Distribution List

Dr. Robert G. Smith
Office of the Chief of Naval Operations, OP987H
Personnel Logistics Plans
Washington, D.C. 20350

Naval Training Equipment Center
ATTN: Technical Library
Orlando, Florida 32813

Human Factors Department
Code N215
Naval Training Equipment Center
Orlando, Florida 32813

Dr. Alfred F. Smode
Training Analysis and Evaluation Group
Naval Training Equipment Center
Code N-00T
Orlando, Florida 32813

Dr. Gary Poock
Operations Research Department
Naval Postgraduate School
Monterey, California 93940

Dean of Research Administration
Naval Postgraduate School
Monterey, California 93940

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
Code RD-1
Washington, D.C. 20380

Mr. Arnold Rubinstein
Naval Material Command
NAVMAT 98T24
Washington, D.C. 20360

Bureau of Naval Personnel
Special Assistant for Research Liaison
PERS-OR
Washington, D.C. 20370

Navy Personnel Research and Development Center
Manned Systems Design, Code 311
San Diego, California 92152

Navy Personnel Research and Development Center
Code 305
San Diego, California 92152

Navy Personnel Research and Development Center
Management Support Department
Code 210
San Diego, California 92152

CDR P. M. Curran
Human Factors Engineering Division
Naval Air Development Center
Warminster, Pennsylvania 18974

Mr. Ronald A. Erickson
Human Factors Branch
Code 3175
Naval Weapons Center
China Lake, California 93555

Dean of the Academic Departments
U.S. Naval Academy
Annapolis, Maryland 21402

Mr. J. Barber
HQS, Department of the Army
DAPE-PBR
Washington, D.C. 20546

ONR Code 455, Technical Reports Distribution List

Dr. Joseph Zeidner
Technical Director
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, Virginia 22333

Director, Organizations and
Systems Research Laboratory
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, Virginia 22333

Dr. Edgar M. Johnson
Organizations and Systems Research
Laboratory
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, Virginia 22333

Technical Director
U.S. Army Human Engineering Labs
Aberdeen Proving Ground, Maryland 21005

U.S. Air Force Office of Scientific
Research
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, D.C. 20332

Dr. Donald A. Topmiller
Chief, Systems Engineering Branch
Human Engineering Division
USAF AMRL/HES
Wright-Patterson AFB, Ohio 45433

Air University Library
Maxwell Air Force Base, Alabama 36112

Defense Documentation Center
Cameron Station, Bldg. 5
Alexandria, Virginia 22314 (12 cys)

Dr. Stephen J. Andriole
Director, Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, Virginia 22209

Dr. Judith Daly
Cybernetics Technology Office
Defense Advanced Research Projects Agency
1400 Wilson Blvd.
Arlington, Virginia 22209

Professor Douglas E. Hunter
Defense Intelligence School
Washington, D.C. 20374

Dr. Richard S. Heuer, Jr.
ORPA/AMERS
Washington, D.C. 20505

Dr. Alphonse Chapanis
Department of Psychology
The Johns Hopkins University
Charles and 34th Streets
Baltimore, Maryland 21218

Dr. Meredith P. Crawford
Department of Engineering Administration
George Washington University
Suite 805
2101 L. Street, N.W.
Washington, D.C. 20037

Dr. Ward Edwards
Director, Social Science Research
Institute
University of Southern California
Los Angeles, California 90007

Dr. Kenneth Hammond
Institute of Behavioral Science
University of Colorado
Room 201
Boulder, Colorado 80309

ONR Code 455, Technical Reports Distribution List

North East London Polytechnic
The Charles Myers Library
Livingstone Road
Stratford
London E15 2LJ
UNITED KINGDOM

Professor Dr. Carl Graf Hoyos
Institute for Psychology
Technical University
8000 Munich
Arcisstr 21
FEDERAL REPUBLIC OF GERMANY

Director, Human Factors Wing
Defense & Civil Institute of
Environment Medicine
Post Office Box 2000
Downsville, Toronto, Ontario
CANADA

Dr. A. D. Baddeley
Director, Applied Psychology Unit
Medical Research Council
15 Chaucer Road
Cambridge, CB2 2EF
UNITED KINGDOM

Dr. J. Baal Schem
Acting Director
Tel - Aviv University
Interdisciplinary Center for Technological
Analysis and Forecasting
Ramat-Aviv, Tel-Aviv
Israel

Dr. Amos Tversky
Department of Psychology
Stanford University
Stanford, California 94305